# **Md Firoz Ansari**

Mobile: +91-9860154411

E-Mail:mdf0288@gmail.com

**Profile Summary:**

* 8.3+ Years of **IT** experience in **Informatica, Informatica Cloud, Hadoop, Talend, Power BI** and **OBIEE.**
* Exposure to onsite-offshore model and client facing engagements.
* Worked at client location (onsite) for requirement gathering.
* Extensive knowledge of ETL, including **Informatica**, **Informatica Cloud** and **Talend.**
* Strong knowledge of Data warehousing, **OLTP** and **OLAP** Concepts.
* Good knowledge of creating **mapping** and **workflows** in Informatica 8.6/9.5/10.01.
* Strong knowledge of using lookup, expression,sorter,rank,filter,sequence generator,union,join etc.
* Extensively worked on Data Extraction, Transformation and Loading with RDBMS,Flat files and XML.
* Also having good knowledge on **Azure SQL**, Azure Data warehousing and **Azure Data Factory**. Developed OLAP Cubes by using **SQL Server Analysis Services (SSAS),** and defined data sources, data source views, Dimensions, Measures, Hierarchies, Attributes, Calculations using multi-dimensional expression (MDX), Perspectives and Roles.
* Integrated Custom Visuals based on business requirements using **Power BI desktop**.
* Utilized **Power BI** to create various analytical dashboard that depicts critical KPIs.
* Worked with **DAX** queries in cubes and Power BI Dashboards
* Experience with building **Metadata Repository** at three layer (Physical Layer,Bmm Layer & Presentation Layer)
* Capable of processing large sets of **structure**, **semi-structured** and **unstructured** data.
* Develop ETL job in Talend to load data from **ASCII**, **EBCDIC**, Flat files to HDFs.
* Solid understanding of the **Hadoop** files distributing system.
* Very good understanding of Partitions, Bucketing concepts in Hive and designed both Managed and External tables in Hive to optimize performance.
* Solid knowledge & Experience on Hadoop architecture, Cloudera, Hive, Impala, Sqoop etc.
* Extensively worked on **SQOOP** to transfer the data between Hadoop and relational database.
* Good knowledge of Dimensional Modeling.
* Good understanding of designing Star and Snowflake Schema.
* Very good understanding of writing SQL and PL/SQL code.

**EDUCATIONALQualifications**

|  |  |  |
| --- | --- | --- |
| **Qualification** | **Name of the School/Institute** | **University/Board** |
| B.E  Year-2011 | PIET,Nagpur | RTMNU |

**EXPERIENCE PROFILE**

|  |  |  |
| --- | --- | --- |
| **Organization** | **Designation** | **Duration** |
| Cloudmoyo India Pvt.Ltd.Pune | Team Lead | 8th Dec 2016 to till date |
| PRGX India Pvt. Ltd.Pune | Senior Developer | 13th Jan 2014 to 2nd Sept 2016 |
| Vcc Technologies Pvt.Ltd,Pune  (A Group Company of Vyomlabs Pvt.Ltd) | Engineer | 23rd July 2012 to 31st Dec 2013 |

TECHNICAL SKILLS

**ETL Tools** : Informatica 8.6/9.5/10.0.1,Informatica Cloud, Talend 5.2, 6.0.1,Talend Big Data

**Reporting Tools** : Microsoft Power BI, OBIEE 11G & SAP BO.

**Other Tools** : SQL Developer, Hue Browser, Putty, TOAD, ISQL\*Plus.

**Operating Systems** : Windows XP, Windows VISTA, Windows10, Redhat Linux (RHEL5).

**Languages** : SQL, PL/SQL, HiveQL, Impala, SQL, Shell Scripting, C, Core Java.

**RDBMS** : MS SQL Azure, DB2, Oracle 11g/10, MySQL, SQL Server, MS Access.

PROJECT DETAILS

1. **Project Title** : Dealer Revenue Share Program

**Client** : **General Motors**

**Team Size** : 10

**Languages** : SQL, UNIX

**Tools** : Informatica Power Center, Informatica Cloud & Power BI

**Operating Systems** : Windows 10

**Environment** : MS SQL Server, MS SQL Azure, DB2

**Description:**

OnStar would like to drive traffic to the Dealer Revenue Share Program Tile by displaying more actionable data to help dealers meet their qualification goal and understand the potential for profits over the next few years. The Dealer Revenue Share Program that contains KPIs, Payment Information, and VIN level details. In addition, OnStar would like to leverage the program page to show data on Sales and Service Consultants to provide a management tool for dealers. Dealer Revenue Share Program Page to include charts, salesperson information, and trend analysis. Power BI for the existing program page expands our dashboard capabilities and offers an opportunity to provide additional content to program users

**Key Roles/Responsibilities and Impact:-**

* Understanding the business rules and sourcing the data from multiple source systems.
* Designed the mappings to extract data from SQL Server, DB2 and flat file sources and load the data into the targets.
* Implemented Type1, Type2 and Incremental load strategies.
* Involved in Staging Area to perform Data Cleansing.
* Used Shared folders to create reusable components (Source, target, email tasks).
* Developing Informatica mappings & tuning them when necessary.
* Used SQL queries to validate the data after loading.
* Published Power BI Reports in the required originations and Made Power BI Dashboards available in Web clients and mobile apps.
* Used Power BI Gateways to keep the dashboards and reports up to date.
* Published reports and dashboards using Power BI.
* Created and managed Source to Target mapping documents for all Facts and Dimension tables.
* Creation of Business test scenarios to ensure Business requirements are met through the project deliverables

**2**. **Project Title** : **RULE 11**

**Client** : **Kansas City Southern Railway (KCS)**

**Team Size** : 9

**Languages** : SQL, Shell Scripting

**Tools** : Informatica Power Center 10.1.0, Informatica Cloud & Power BI

**Operating Systems** : Linux, Windows 10

**Environment** : MS SQL Server, MS SQL Azure, DB2

**Description:**

KCS has an existing system called Financial Data Warehouse which deals with waybill (goods being carried on a vehicle) information. Waybill Fact information connected to several dimensions like customer, equipment, location, road, shipment, user, vendor etc. which provides necessary information to track relationship in between.Rule-11 comes into picture when we need to populate billing information with multiple segments.

**Key Roles/Responsibilities and Impact:-**

* Creation of Landing, Stage and Base objects required for the data load
* Creation of Landing to Stage mappings and setting of foreign key relationships
* Setting up of the Trust percentage for the required fields.
* Coordinating with the business analyst to get the clear understanding of the business and meet their expectation.
* Created and configured Sessions in Informatica workflow Manager for loading data into tables from various heterogeneous database sources like Flat Files, relational Data bases.
* Created Mappings for Slowly Changing Dimension tables.
* Used the version control to maintain the mapping versions.
* Setting up of validation rules and creation of Batch groups
* Creation of Business test scenarios to ensure Business requirements are met through the project deliverables
* Preparation of requirement and technical design document.
* Ensuring to focus on deadlines and deliverables.

**3**. **Project Title** : **Car Hire**

**Client** : **Kansas City Southern Railway (KCS)**

**Team Size** : 7

**Languages** : SQL, Shell Scripting

**Tools** : Informatica Power Center 10.1.0, Informatica Cloud & Power BI

**Operating Systems** : Linux, Windows 10

**Environment** : MS SQL Server, MS SQL Azure, DB2

**Description:**

This Project was developed for KCS, in order to manage the Car hire related information, which involves different cars hired to deliver the goods for KCS.

Objective of this project was to retrieve Car hire related data from different sources like proprietary flat files and DB2 database system .Enterprise level data warehouse was delivered by CloudMoyo, where we have maintained 48 months of data which stands for latest 4 years rolling period in fact table and different dashboards and KPI were delivered using Power BI

**Key Roles/Responsibilities and Impact:-**

* Getting requirements from business and develop it.
* Involved in Staging Area to perform Data Cleansing.
* Build the Data Model according to the requirements of the reports.
* Involved in building the mappings and workflows to populate the target warehouse.
* Created multiple transformations required for the mappings.
* Developed Mapping to load the data in slowly changing dimension.
* Weekly/Monthly operation that involves ETL operations and Report Testing.
* Used the version control to maintain the mapping versions.
* Co-ordinate with the QA team in system testing.
* Provided UAT and Production support for any issue.

**4. Project Title** : **HiPer**

**Client**  : **Amazon**

**Team Size** : 8

**Languages** : HiveQL, Impala, SQL, Shell Scripting

**Tools** : Talend, Hue Browser, Putty

**Operating Systems :** Linux, Windows 7

**Environment** : Big Data Hadoop, MS SQL Server, MySql.

**Description**:

HiPer stands for High Performance Computing. This project is started with the vision to optimize the current process from fetching the data from client and providing it to Audit tools by using Hadoop. This involves migrating existing processes which are currently using AS400, MS SQL Server, Informatica and few other tools into Hadoop, Hive, Impala and Talend.

**Key Roles/Responsibilities and Impact:-**

* Design Templates for Requirement documents.
* Load and transform data into HDFS from large set of structured data using Talend Big data studio.
* Creating Hive Scripts, Identifying parameters as per requirement to apply transformation and perform Unit testing on data as per design.
* Conduct proof of concept (POC) in Hadoop as per requirement.
* Design, develop, unit test, scripts for data Items using Hive script and Export final tables from HDFS to SQL server using SQOOP.
* Design Templates for Parameter and sequence file.
* Designed and developed Reusable Components
  + The Reusable components help to automate different processes in the project. These components involve Master Script, Compute Stats, Space Usage Report, Data Archival package.
* Implement the Logical and Physical data model for Meta Data Repository(MDR)
  + Meta-data repository is a set of database table which does the task of configuring the jobs that perform the task of data-loads and data-builds. It is also now been used for different purposes such as load balancing and performance reporting tool.
* Generate Builds using HiveQL
  + Created Data Builds using HiveQL which involved writing Hive scripts to merge, dedup join RAW Data received from client and perform the operations defined in the requirement documents.
* Lead the team for few individual modules.

**5. Project : Talend Automation**

**Key Roles/Responsibilities and Impact:-**

* Create Mappings from source to staging and staging to data-warehouse.
* To make the suggested model work in an automated fashion with a single button click which impacted PRGX in reducing the people based dependency and delivering the data at a faster pace.

Designing and Implementing MDR: Meta-data repository is a set of database table which does the task of configuring the jobs that perform the task of data-loads. It is also now been used for different purposes such as load balancing and performance reporting tool.

* To create an MDR that would work for over 80 clients required a thorough study of the types of the files received and to make provisions in the MDR for each of the file type per client per schema.
* The extensive research and development resulted in an MDR that is been used as single point of truth for loading the data and answering the Management various questions.

Automate QA Process:

* The QA process is comparing output of legacy tool and new Talend tool. For which we had to put lots of Manual efforts. Which we have reduced to 90 percent by automating this process.
* To automate the QA process impacted in the delivery model to be faster and less prone to human errors.

**6. Project:**

**Project Title : Walmart AP Vendor DataMart**

**Client : Walmart, USA.**

**Team Size :** 10

**Environment :** Informatica Power Center 8.6, MS SQL Server 2005

**Description :**

Wal-Mart a global leader in Retail sector wants to have reporting DataMart for Accounts Payable and Vendor data to audit their claims. We are processing the flat files data received through AS 400 systems – FTP File Servers to MS SQL Servers in staging databases of ODS. After this data moves to DataMart dimensions which is also used for Recovery Audit and Analysis purpose. Data comes in zipped and/or encrypted format.

* Pre load activities like delimiter stats, line stats, data issues resolution.
* Transfer data to MS SQL Server database using Informatica mapping customized as per client to client.
* After loading the data, unit test the data, generate the report statistics for the same and sent it to QC.
* Upon getting clearance from QC team send the completion details to BIA onsite.
* Lookup transformations to compare the source and target values and Update strategy transformations are used for Update or Delete the records
* Created test cases, defect logs for mappings and workflows.

ACHIEVEMENTS

* Passed the exam for “**AZ-900 Microsoft Azure Fundamentals**”
* Our team earned “**Circle of Excellence**” Award for one year 2017 in CLOUDMOYO.
* Awarded with the Employee of the month for my flexible timing, extended work and ownership during the ramp down phase of the project.
* Earned SPARKATHON Award in PRGX - 2015.
* Our team earned “Circle of Excellence” Award for two consecutive years 2014 and 2015 in PRGX.
* Cloudera Certified Data Analyst: Data Analyst Training using hive, pig and impala with Hadoop

PASSPORT & VISA DETAILS

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Name as on**  **Passport** | **Passport Number** | **Date of Issue** | **Expiry Date** | **Place of Issue** | **US VISA** |
| Md Firoz Ansari | K1333893 | 04/11/2011 | 03/11/2021 | Patna | B1/B2 |

I hereby declare that the above mentioned information is correct to the best of my knowledge and I bear the responsibility for the correctness of the above mentioned particulars.

Place: Pune

Date : (Md Firoz Ansari)